Collection and Analysis of Multi-dimensional Network Data for Opportunistic Networking Research

Theus Hossmann[a,*], George Nomikos[a], Thrasyvoulos Spyropoulos[b], Franck Legendre[a]

[a] *Communication Systems Group, ETH Zurich, Switzerland, lastname@tik.ee.ethz.ch*
[b] *Mobile Communications, EURECOM, France, firstname.lastname@eurecom.fr*

Abstract

Opportunistic networks use human mobility and consequent wireless contacts between mobile devices to disseminate data in a peer-to-peer manner. Designing appropriate algorithms and protocols for such networks is challenging as it requires understanding patterns of (1) mobility (who meets whom), (2) social relations (who knows whom) and (3), communication (who communicates with whom). To date, apart from few small test setups, there are no operational opportunistic networks where measurements could reveal the complex correlation of these features of human relationships. Hence, opportunistic networking research is largely based on insights from measurements of either contacts, social networks, or communication, but not all three combined.

In this paper we analyze two datasets comprising social, mobility and communication ties. The first dataset we have collected with Stumbl, a Facebook application that lets participating users report their daily face-to-face meetings with other Facebook friends. It also logs user interactions on Facebook (e.g. comments, wall posts, likes). For the second dataset, we use data from two online social networks (Twitter and Gowalla) *on the same set of nodes* to infer social, communication and mobility ties. We look at the interplay of the different dimensions of relationships on a pairwise level and analyze how the network structures compare to each other.

Keywords: Opportunistic Networks; Human Mobility; Online Social Networks; Facebook; Twitter; Gowalla; Complex Networks; Multi-dimensional Network Analysis

1. Introduction

The rapid proliferation of small wireless devices (e.g., smart phones) creates ample opportunity for novel applications [1], as well as for extending the realm of existing ones [2, 3]. *Opportunistic networking* is a new networking paradigm that is envisioned to complement and extend existing wireless infrastructure such as 3G and WiFi: Mo-

*Corresponding author

bile devices exploit communication opportunities by exchanging data whenever they are within mutual wireless transmission range (*in contact*).

Algorithms and protocols (e.g., routing protocols) for opportunistic networks were originally largely based on random decisions [4], not accounting for heterogeneity in terms of capabilities of devices and behavior of people carrying them. Such random protocols typically require large amount of resources for timely delivery of content (e.g., epidemic spreading of messages). To overcome this, more recent protocols exploit node heterogeneity in order to make educated decisions to provide good performance at limited resource usage. Examples are routing protocols exploiting structure in social ties [5, 6, 7] or structure in mobility ties [8, 9]. Simulations show that efficiency is much better than for random protocols.

Designing and analyzing efficient protocols is challenging, as it requires knowledge about various aspect of human behavior. Relevant questions are: *Which nodes have frequent contacts and hence are good relays? Which nodes are socially related and hence trust each other and are willing to cooperate? Which nodes communicate with each other and need fast routes between them?* In fact, we can assume that these three dimensions of social, communication and contact relations are correlated at least to a certain degree. However, it is largely unknown how strong this correlation is, how it can be exploited for opportunistic networking and how it affects performance of existing protocols.

To date, there are only few small deployments of opportunistic networks [10, 11, 2, 3] from which practical insights into the correlation of social, mobility and communication ties could be gained. Hence, research in this direction is largely based on insights from empirical analysis of datasets typically capturing only one or two of the aspects of relations, but not all three combined.

Example datasets are *mobility traces* (some of which also contain information about social ties between the nodes) from WLAN Access Point associations [12] or Bluetooth contacts [13, 14, 15]. Analysis of such traces has shown that there is some correlation of mobility and social connections [13, 15]. However, these analyses do not consider which nodes would actually actively *communicate* and interact with each other in an opportunistic application (i.e., who is interested in content of whom, who sends messages to whom). To also capture this aspect, we want to collect datasets comprising all *three* dimensions.

While mobility and social connections can be measured, the question of who communicates with whom using opportunistic applications is difficult to answer, as there are only few – and mostly small – deployments of opportunistic applications [3, 10, 11]. However, we assume that opportunistic applications are of social nature and we speculate that they would create similar communication patterns like today's online social network and Web 2.0 platforms, such as Facebook or Twitter. In fact, current online social networks could be run over opportunistic networks [2, 3].

Facebook is a typical, and to date the most widely used, representative of an online social networking service, fostering communication and distribution of (user generated) content among friends. It provides an API for application development, allowing us to create an application – called Stumbl – to measure all three dimensions of interest. Using the Facebook API, Stumbl records communication and social ties of its users. Additionally, it asks participants to report their meeting data regularly, to also cover the

mobility dimension of users' relations (i.e., how often, how long and in what context users meet their Facebook friends).

While the data we have collected with Stumbl have a great level of detail, it is limited to a relatively small set of users. Thus, we use a second dataset to corroborate our findings. As a second source of data we use the two social networks Gowalla and Twitter. The mobile social network Gowalla[1] lets users check in to close by spots (e.g., restaurants, office buildings, home, etc.) using an application and the location services provided by smart phones. Further, Gowalla users can connect their accounts to their respective Twitter account, which allows to collect data about whom they communicate with on Twitter. This allows us to gather data from a larger number of users, although sparser than the Stumbl data.

Together, these datasets provide rich information about the different aspects of human relationships. The contributions and structure of this paper can be summarized as follows.

1. We describe the two datasets and methodologies to collect self-reported data of user behavior (Sec. 3 and Sec. 4).

2. We analyze the two datasets with a focus on the interplay of ties across dimensions: how do the characteristics of a tie in one dimension affect the properties in another dimension (Sec. 5).

3. Diving deeper, we perform a multi-dimensional network analysis, comparing the structure ("hubs", communities, small-world property, etc.) of the graphs across all three dimension (social, meeting and communication) (Sec. 6).

4. Further, we discuss implications of these results for opportunistic routing and traffic modeling (Sec. 5 and 6).

2. Related Work

In this section we discuss advantages and disadvantages of some well known sources of mobility, social and communication data, and give a brief overview of the insights gained from empirical analyses. Note that the analysis of such data is an interdisciplinary effort (ranging from social sciences, to epidemiology, to urban planning and mobile networking). Consequently, there exists a big body of related literature – even if many of the larger data sources (e.g., mobile phone data, online social networks) have only existed for few years. We try to focus on few studies which we consider the most relevant to opportunistic networking.

Typically, wireless contacts between mobile devices are measured via Bluetooth [15, 14] or WLAN ad hoc [16]. Contact traces measured in experiments have proven very fruitful for analyzing pairwise contact and inter-contact patterns. The debate is still on-going whether inter-contact times are heavy tailed [17], have an exponential cut-off [18] or differ from pair to pair [19]. Different studies have related contact patterns to social information [13, 15], finding that social ties heavily influence contact patterns.

[1] Gowalla launched in 2009 and closed in 2012.

Collecting contact traces with Bluetooth or WLAN ad hoc has the advantage that contacts between devices can be measured directly, but comes at a high cost: experiments are complex and expensive and hence usually limited to a small number of participants.

Indirectly, contact information can also be estimated from shared location patterns. In this direction different studies have analyzed WLAN access point associations and found temporal (i.e., diurnal and weekly patterns) and spatial (i.e., frequently visited hotspots) regularity of human mobility [20, 12, 21]. More recently, studies have confirmed such regularities based on larger datasets (with hundreds of thousands of users) using mobile phone location data from mobile network operators [22]. These studies imply a rather high predictability of human mobility [23]. Connecting mobile phone location data with communication ties (who calls/texts whom), the authors of [24] find that mobility is a good predictor for communication. I.e., if two nodes manifest similar mobility patterns, the chances of communication is considerably higher than for nodes with different mobility patterns.

On one hand, analysis of mobile phone data has big advantages over other sources of data: they allow studying huge number of nodes which represent a rather unbiased sample of the population. On the other hand, there are drawbacks in terms of precision: the location data is very coarse grained since location is inferred from the location of the base station a phone is associated with. Further, social data is typically very limited to information about who calls or text whom.

A third source of data is online social networks. In sociology, online social networks and their relation to offline friendships has been studied for a long time (e.g., [25]). Recent studies also incorporate mobility data available in social networks to quantify the correlation between social ties and mobility (i.e., location reports from users' checkins) [26, 27]. Using online social networks to collect data has several advantages: Social information and data about communication (at least the communication that happens within the platform of the OSN) between users is typically readily available either publicly or upon permission by the user. Further, the number of users can be very high. However, the drawback of data from OSNs is that mobility data is sparse and limited to the occasional check in of users.

In this direction, we explore here an option to make location data less sparse: to use OSNs for making surveys asking people to report their mobility. This paper extends previously presented work [28] by providing an additional analysis of the *structure* of the networks formed from the various tie types. Further, we add a second dataset of mobility, social and communication ties, which contains a larger set of nodes, but data is sparser.

3. Stumbl Application and Dataset

To measure contacts, social ties and communication, we have implemented Stumbl as a Facebook application. In this section we briefly discuss the Stumbl application (3.1) as well as the Stumbl experiment and resulting dataset (3.2). Finally, we also discuss limitations of the methodology and collected data (3.3). A more detailed description of the application and experiment can be found in [29].

3.1. The Stumbl Application

Facebook provides an API for authorized (by the user) applications to access user data. Our Stumbl application[2] uses this API to retrieve the user's social connections and Facebook communication events. Additionally, we ask the users to regularly report whom of their friends they meet face-to-face, by filling in a survey form in the Stumbl application. One big benefit of integrating Stumbl as an application in the Facebook website is that it is a convenient way for many people to report their meeting data: Since visiting the Facebook website is part of the daily routine for many people, the barrier to fill in the survey is small.

When a user joins the Stumbl experiment, there are two main phases of participation.

Initialization Phase: In a one time initialization step, the user is asked to select a subset of her Facebook friends which she meets face-to-face regularly (at least once a month). We will refer to this subset of Facebook friends as the *Stumbl friends*. The reason for selecting a subset of the friends for the survey is two-fold. First, most users have large number of Facebook friends, many of which living far away. These pairs typically have only very rare meetings (weak ties). In order to keep the effort for reporting data as small as possible, we wanted to exclude them from the input interface. Second, we are mainly interested in the meeting patterns of people who see each other frequently (strong ties), as such meetings are more predictable than the random occasional meetings[3].

In our experiment we have limited the number of Stumbl friends to 20. Typically, a user regularly sees less than 20 of her Facebook contacts, as we will report later. The selection of 20 friends hence does not narrow the data we gather. Note that the users have the option to change their selection of Stumbl friends during the experiment.

To complete the initialization step, Stumbl asks the user to classify the relationship type to each of the Stumbl friends as one or more of *family, friend, colleague* or *acquaintance*. As "friendship" on Facebook is a very broad term characterizing a wide range of actual social relationships, we use this classification for a more fine tuned analysis of the social dimension of relations. Note, however, that this classification of the tie *type* is not necessarily a good indicator for tie *strength* [30]. Social tie strength is not currently measured by the Stumbl application.

Reporting Phase: After the initialization step follows the recurring report of face-to-face meetings. As automated measuring of face-to-face meetings typically requires special equipment (iMotes [14] or phones equipped with special software [15]) and is costly and complex, we rely on self-reported data to assess the mobility dimension of relations. Correlating self-reported and measured (via Bluetooth) proximity has shown that the quality of self-reported proximity data drops when reporting events more than seven days back in time [15]. To ensure a good level of accuracy for the reported information, we choose a reporting interval of one day: The Stumbl users are

[2] `http://apps.facebook.com/stumbl_app/`

[3] Note that the occasional random meetings of weak ties can be very beneficial, for instance for opportunistic routing protocols as "short cuts". However, they are typically not predictable and protocols can not rely on them. Decisions have to be made depending on strong and predictable mobility ties.

Figure 1: Stumbl screenshot. For each Stumbl friend, context, number and total duration of meetings can be reported (for the previous day). Options are chosen to capture a range of different meeting behaviors.

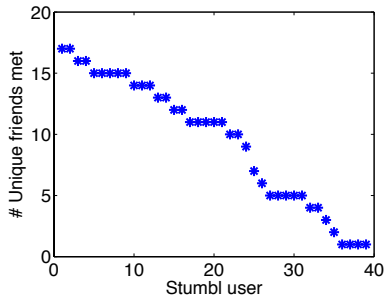
asked every day (reminded by E-Mail) to visit the Stumbl application and fill in the questionnaire about whom of their Stumbl friends they met the previous day[4]. Thus, the collected data has a temporal resolution of meetings of one day.

For each friend a user reports a meeting, additional information has to be provided about (i) how often she saw the friend (options are 1, 2-3, 4-5, more times), (ii) for how long in total these meetings lasted (with options 0-10 min, 10-30 min, 30 min - 1 hour, 1-2 hours, more than 2 hours), and (iii) the contexts of the meetings (given the options *work*, *fun*, *home*, *meal*, *other* for selection). These additional features allow us to make a more fine grained analysis of the contact data.

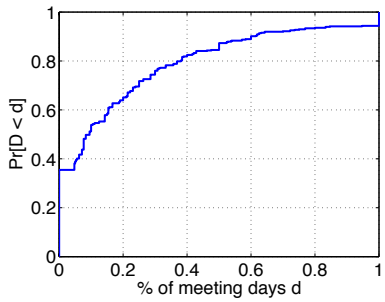
Fig. 1 shows the input interface as participants see it. We designed the interface such that we can collect a maximal amount of data with as small an effort as possible by the user. From experience and user reports, we know that the input requires less than 5 Minutes per day, a target we had set to motivate daily participation.

In order to capture communication between a user and her Stumbl friends, the application uses the Facebook API to query for interaction events, every time meeting data is submitted. We collect the following three types of interaction to which the API

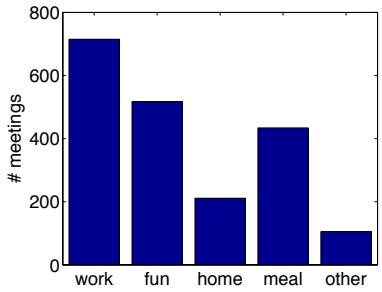
[4]Note that with the check-in service *Places*, Facebook also provides a platform for recording user location and meetings (tagging people at the same location). However, this would require users to check-in and tag people at every meeting and is too cumbersome to ask. Also, since check-ins and tags show up in the user profile, this methodology of recording meetings would have serious privacy issues.



(a) Ranked numbers of unique friends met by Stumbl users.



(b) CDF of percentage of days the pairs meet during experiment.



(c) Number of meetings reported per context.

Figure 2: Overview of Stumbl meeting statistics.

provides access[5]. *Wall posts:* Users post content (messages, photos, videos, links, etc.) on each others wall. *Comments:* Wall posts can be commented on. *Likes:* As a sign of approval, any item on the wall can be "liked".

These communication events are time stamped. They are all directed (e.g., a user writes on another user's wall), and we collect both incoming and outgoing events.

Summarizing, Stumbl records social ties (friend, family, colleague, acquaintance), Facebook communication (wall posts, likes, comments, tags) and meeting data (number, duration and context of meetings) which allows us to get insight in three dimensions of the relationships of a Stumbl user.

3.2. The Stumbl Experiment

In this paper, we report results from a preliminary experiment using the Stumbl application, which we used to gain experience with application and user behavior – and which also led to a first interesting (but limited in size) dataset. The experiment ran for three weeks between August 16th 2010 and September 6th 2010. At the beginning

[5]A fourth communication mechanism, private messages, is not accessible by the API for obvious privacy reasons.

	Posts	Comments	Likes	Total
Nr. of Events	199	341	103	643

Table 1: Total number of registered Facebook communication events between Stumbl friends per event type.

of the experiment, we recruited participants mainly by personal invitations, which led to a total of 39 users providing useful information. In order to provide incentives for these users to persistently report their meeting data during the experiment, the users participated in a raffle. To provide the right incentives, the chances of winning were dependent on two factors: the number of days the application was visited, and the number of their friends who registered as Stumbl users. While these raffles should be incentives to provide data regularly, they should not provide incentives to provide false data. In the following, we provide an overview of the dataset we collected during this experiment.

During the 21 days of the experiment, on average 22 of the 39 participants reported meeting data. This means that users were quite persistent in participating and shows that the incentives for participating regularly worked well. We will now report some general statistics about the collected data to provide a general impression of the dataset.

On average, users selected 14 *Stumbl friends* in the initialization step. 11 users selected the maximum allowed 20 users. The number of Stumbl friends the user actually reported meetings with throughout the experiment is lower than the number of Stumbl friends, as shown in Fig. 2a. On average a user reported meeting 9.5 unique Stumbl friends during the experiment. The maximum is at 17 unique friends and hence lower than the 20 allowed. We conclude that the selection of 20 friends does not narrow the number of pairs for which we receive meeting reports.

In total, we have 498 pairs of Facebook users[6] in our Stumbl dataset. Fig. 2b shows the cumulative distribution function of how often these pairs met. As users did not report their meetings every day, we divide the number of days a pair meets by the number of days we have self-reported meeting data for the given pair (including days where they report no meeting). Thus, the figure shows the percentage of days the pairs met. Roughly 65% of the pairs met at least once and almost 5% of pairs report meeting every day.

Further, we want to analyze the contexts (*work*, *home*, *fun*, *meal*, *other*) of the meetings. Fig. 2c shows how the meetings are split among the different contexts. We observe that most meetings happened at work, but also for the other contexts we have quite large numbers of meetings reported.

Next, we want to provide an overview of the social tie types our measurements cover. Fig. 3 shows how the 498 Stumbl pairs are divided into *family*, *friends*, *colleagues* and *acquaintances*. We observe that most of the pairs are classified as friend

[6]For 47 of these pairs, we have mutual meeting reports data, i.e, both nodes participate in the Stumbl experiments. For the rest only one node reported data.

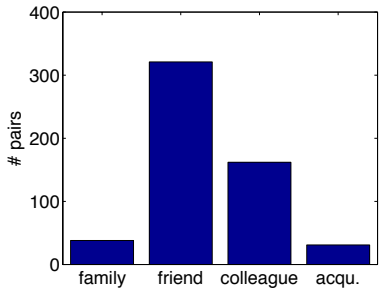


Figure 3: Number of pairs per social tie type.

or colleague. Relatively few pairs are of types family or acquaintance. Note that the user can specify more than one type of social tie per Stumbl friend. Hence, the number of pairs per type sum up to more than 498.

In terms of communication events, Tab. 1 summarizes the number of events we recorded during the experiment. With a total number of 643 communication events, we have a large enough sample to provide statistics about communication. In total, we saw communication between 91 or 18% of the 498 pairs.

These statistics give a separate overview about each of the three dimensions of relationship we measure. In Sec. 5 we will analyze how the different aspects of relationship correlate with each other. We will look at questions like: *How does the type of social tie affect meeting probabilities and communication probabilities? How do meetings relate to probabilities of communicating?*

3.3. Limitations and Validation of Dataset

We now want to adress potential bias and limitations of our dataset and the methodology of collecting self-reported meeting data using a Facebook application.

The 39 users have an average of 252 friends in their Facebook social graph. This is considerably more than the average friend count of 130 reported by Facebook[7]. We assume that the large number of Facebook friends does not mean that the average Stumbl user is more sociable than an average person. Rather, it means that the Stumbl users are more active Facebook users. While this may cause a bias in the measurement, we believe that the Stumbl users may actually be more representative users of opportunistic networks, as we expect them to be well-versed users of new technologies.

As Stumbl users were recruited based on personal invitations by the authors of this study and by word-of-mouth recommendation, the Stumbl users present a rather local group of people (most are researchers or students living in few cities). In the future, we plan to extend Stumbl and use it for experiments with broader audience.

Another concern is that the self-reported meeting data may be erroneous because the user does not recall meetings correctly or decides to provide wrong information. In

[7]Facebook Statistics: `http://www.facebook.com/press/info.php?statistics`

order to estimate the severity of these effects, we validate the data where possible. We do so by looking at the 47 pairs of users for which we have mutual meeting data. We find that in 86% of the cases the reports whether or not there was a meeting between a pair matches (i.e., both Stumbl users report that there was a meeting or both report there was no meeting). This seems a quite good correlation. For the cases where both report that there was a meeting, we further check whether their reported meeting counts, meeting duration and meeting contexts match. We find that this is the case in 57% of meeting counts, 66% of durations and 87% of contexts. While not perfect correlation, we conclude that the reports are accurate enough, especially those of meetings or not on a given day and the context in which the meetings happen.

A limitation inherent to the methodology of self-reported mobility data is that Stumbl can only capture meetings between friends. Random encounters of strangers or meetings between familiar strangers cannot be recorded. Thus, on one hand we are limited to the analysis of properties of *strong* mobility ties. On the other hand, Stumbl provides very faceted information for these strong ties, allowing us to make very detailed analyses of the strong *backbone* of opportunistic networks. Note that for analyzing contacts this limitation can be an advantage: Typically, in automatically recorded contact traces, it is hard to distinguish strong and weak ties and it is not a priori clear if a contact is a random encounter or part of a more "meaningful" mobility tie.

4. Gowalla and Twitter Dataset

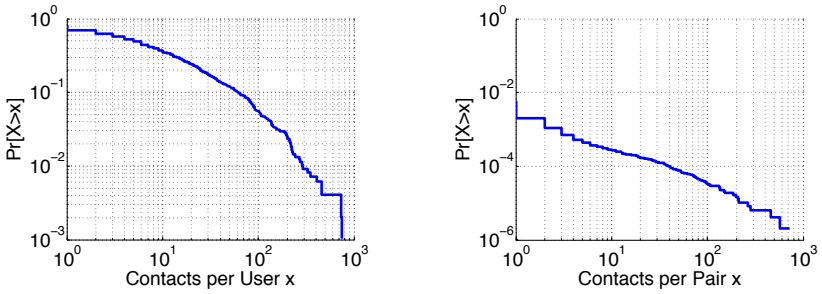
While the Stumbl dataset gives us a great level of detail in information about relationships, it is very limited in size. Thus, to corroborate the findings of our analysis, we use a second dataset, which is more sparse but comprises a much larger population.

We use publicly available data from two online social networks: the geo-social network Gowalla and the micro-blogging network Twitter. Both provide APIs for querying their users' data which we use to collect the datasets. With the Gowalla API we can query the Twitter username and user ID of a Gowalla user (only for the users which provide this information to Gowalla and have their user profile publicly viewable), *which allows us to collect data from both networks on the same set of users, for a large number of users*. In the following, we describe the data collection (Sec. 4.1) and characterize the resulting dataset (Sec. 4.2).

4.1. Data collection

The Gowalla application lets its users check in to close-by spots (e.g., restaurants, office buildings, home, etc.) using an application for smart phones. Such a checkin logs the time, position and context of a user. The location of the user is determined from the GPS of the smart phone. Additionally, Gowalla was an online social network where users maintained a list of their friends.

Using the Gowalla API, third party applications could query the database for spots,



(a) CCDF contacts per user.

(b) CCDF contacts per pair.

Figure 4: Contact statistics (Gowalla dataset).

users, history of checkins, etc[8]. Using this API, we crawled the Gowalla database during a period of 2 weeks in October 2010.

From this crawling, we obtained a dataset of $\sim 470'000$ users, with a total of $\sim 17'000'000$ checkins to $\sim 1'700'000$ different spots globally. From this, we restrict the checkin data which we analyze to the period of 6 months from April to September 2010 and take a subset of users which fulfill the following criteria: i) Their Gowalla data (including checkin history and list of friends) must be public. Further, they must have connected their Gowalla identity to their Twitter account, and their Tweets must be public as well. ii) They must be heavy users of both Gowalla and Twitter. We define a heavy user as a user which sends Tweets[9] and checks in at least 5 out of 7 days. After this pre-processing, we obtain a dataset of 978 users.

Using the Twitter API, we have crawled the followers and Tweets (posts or messages of a maximum of 140 characters length) of these users at the end of October 2010. We get a total of $\sim 1'000'000$ Tweets (like for the checkins, we only consider Tweets from the the period of April to September 2010).

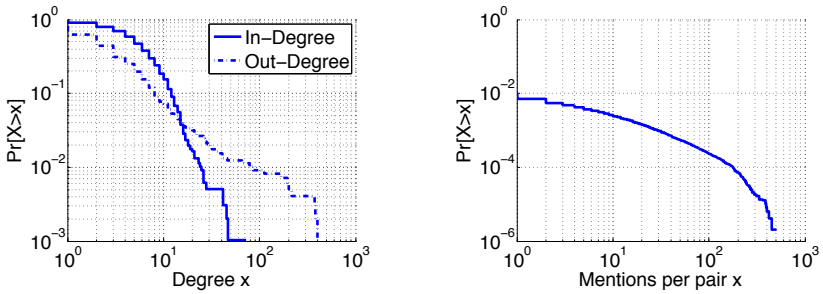
4.2. Dataset description

We will now describe how we infer social, mobility and communication ties from the crawled data.

Social Ties: We have two sources of information about social ties: Gowalla friendships and Twitter follower relationships. The Twitter follower graph is often argued not to be a classical social network but rather a network of *interest*, i.e., users not only follow their friends, colleagues and family, but to a large extent also other people and organizations they are interested in (celebrities, news sources, etc.). Thus, we use the Gowalla social network for our analysis.

[8]User data and especially checkins can be hidden from the public (i.e., made accessible only to the user and his friends) in the privacy settings.

[9]Gowalla allows to automatically publish a Tweet for every checkin. In order to not count users as Twitter heavy users who only Tweet their checkins but do not use Twitter aside from this, we filter out all Gowalla generated Tweets in a pre-processing step.



(a) CCDF communication peers.

(b) CCDF mentions per pair.

Figure 5: Communication statistics.

Mobility Ties: To infer the strength of mobility tie between two users, we want to know how strongly their mobility patterns correlate in space and time, i.e., how often they visit the same place at the same time. Such co-location (or *contacts*) can occur intentionally or can be merely random co-locations because of similar mobility patterns. We define a *contact* between two users, if they check in at roughly the same time at the same spot. As a threshold, we consider two users as collocated if the time difference between their checkins is less than one hour[10].

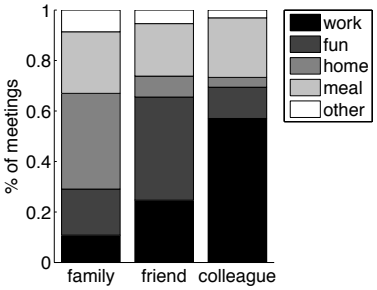
This gives us a total number of $\sim 12'000$ contacts between $2'864$ pairs of users. Fig. 4a shows the CCDF of contacts per user. Fig. 4b shows the CCDF of number of contacts per pair. We observe that both, individual users and pairs show large heterogeneity in number of contacts.

Communication Ties: In order to infer how often two nodes communicate with each other, we account for mentions (i.e., Tweets that address a user by using the @username notation) in the Twitter dataset.

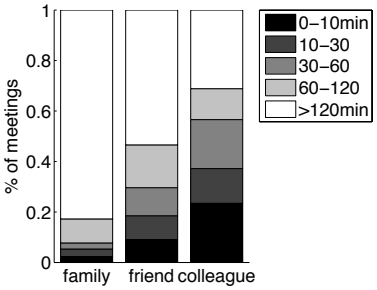
We have a total number of $\sim 37'000$ mentions among $3'787$ pairs of heavy users. Fig. 5a shows the CCDFs of the number of communication peers (i.e., how many peers mention a user and how many peers a users is mentioned by). We see that heterogeneity is higher in the out degree than for the in degree, i.e., there are users mentioning more than 100 different users but no user gets mentioned by 100 other users. Further, Fig. 5b shows the distribution of the number of communication events per pair. Again, we observe that there are highly active pairs with several hundreds of mentions.

5. Social Ties vs. Meetings vs. Communication

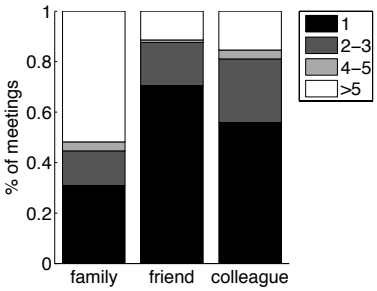
After providing an overview of the datasets, we now present an empirical analysis of how social ties, meetings and communication relate to each other. We also discuss the hints our findings provide for opportunistic routing and traffic modeling.



(a) Tie types vs. context of meetings.

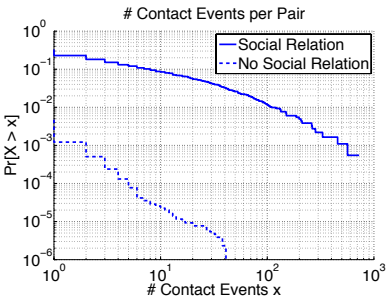


(b) Tie types vs. duration of meetings.

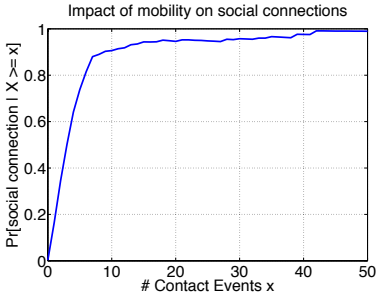


(c) Tie types vs. frequency of meetings.

Figure 6: Dependence of meeting patterns on social tie type (Stumbl dataset).



(a) CCDF number of contacts.



(b) Impact of mobility on social tie probability.

Figure 7: Relation of social ties and mobility ties (Gowalla dataset).

5.1. Social Ties vs. Meetings

First, we look at the relation of social ties and mobility ties. In the Stumbl dataset, where we have information about the *type* of social tie (family, friends, colleagues,

[10]Different threshold values give qualitatively similar results.

acquaintances), we look at how this impacts the meeting behavior. From experience and intuition about human mobility, we expect that meeting patterns of colleagues, family, friends and acquaintances have different characteristics in terms of context, frequency and duration.

As a sanity check, we start by looking at social tie type and meeting contexts. Naturally, we expect the tie type to influence the context of meetings: We meet colleagues at work, family at home, friends for fun, etc. Fig. 6a confirms this by showing the percentage of meetings happening in a given context, split by social tie type[11].

Fig. 6b and 6c show how long and how often pairs meet per day (given that they meet at least once that day). We observe that meetings between family are generally long and frequent. Between friends, meetings are still quite long but typically only once per day. For colleagues, meetings are generally shorter. Such short meetings of colleagues may be just crossing each other, talking briefly or drinking a short coffee during breaks.

In the Gowalla dataset, where we also observe meetings of strangers, we want to know how having a social relation impacts the probability of meeting and vice versa. Fig. 7 shows for the Gowalla dataset[12] that – as expected by intuition – there is very strong dependence of the two dimensions. In fact, Fig. 7a shows that having a social connection increases the probability of having at least one meeting by a factor of more than 100. Further, we can observe in Fig. 7b that people with frequent contacts almost always have a social tie: about 90% of pairs with 7 or more contacts have a social connection.

Summarizing, we find that social ties and mobility ties are closely related. Further, we observe the social tie type has very strong impact on meeting characteristics in terms of context, duration and frequency of meetings. These results are not surprising. Yet, they have implications for example for DTN routing protocols where routing decisions are based on social networks [5, 6, 7]. If the *type* of social link is known to such protocols, this might be useful information, without necessarily having to sample actual contact times. Different conclusions and strategies may be applicable to different tie types: Typically, a tie with frequent meetings is a good carrier in terms of short delivery delay. However, if the frequent meetings are short, the capacity of the contacts may be too small to deliver a large amount of data. For large data transfers, long meetings may be more desirable.

5.2. Social Ties vs. Communication

In the Stumbl dataset, we investigate how the social tie type is related to communication patterns. Fig. 8 reports the average number of communication events per pair during the experiment, split by social tie type. We note that friends and family are the most communicative. Colleagues communicate much less and for acquaintances we find an average of merely 0.3 communication events per pair, not even one fifth of the

[11] We do not show acquaintance relationships here since we observe too few meetings between acquaintances in the dataset to make reliable statements.

[12] For the Stumbl dataset we do not have information about random meetings of strangers. Thus, we can not do the same analysis there.

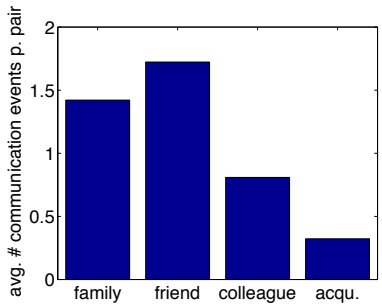


Figure 8: Tie types vs. communication events (Stumbl dataset).

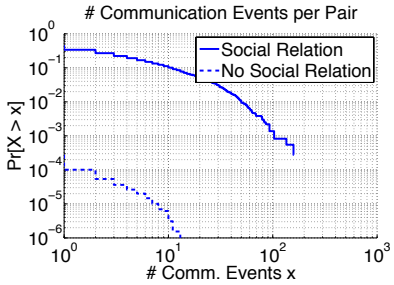


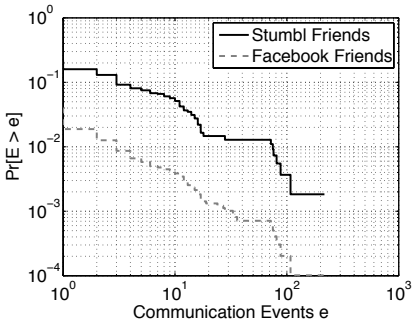
Figure 9: CCDF number of communication events (Gowalla dataset).

communication events an average friend pair shows. Thus, not all nodes with social ties communicate with the same frequency. Instead, communication, or traffic, between pairs of nodes depend on their type of social tie. This is something to consider when simulating opportunistic network traffic. Realistic traffic models should incorporate heterogeneity of social ties and how this reflects in communication patterns.

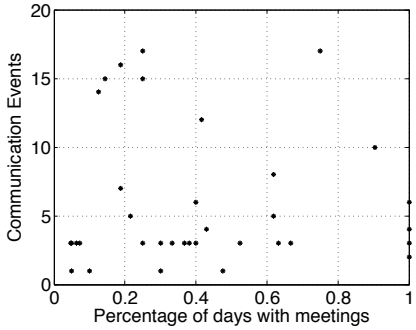
In the Gowalla dataset, we can also measure communication between pairs without social tie. Fig. 9 shows the CCDF of the number of communication events, for pairs with and without a social tie. As expected, pairs with social connection are much more likely to communicate – the figures show that the difference is more than three orders of magnitude. For opportunistic networks, this implies that "fast" opportunistic routes must be mainly established between socially connected pairs.

5.3. Meetings vs. Communication

Finally, we are interested in how contacts affect the probability of communicating and vice versa. *Are we more likely or less likely to communicate with friends to whom we have strong mobility ties? In other words, do we communicate with friends we see face-to-face (e.g., to discuss common experiences) or with remote friends (e.g., to stay in touch)?* To answer this, we compare the number of communication events of Stumbl friends (as representatives of friends to whom we have strong mobility ties) to the

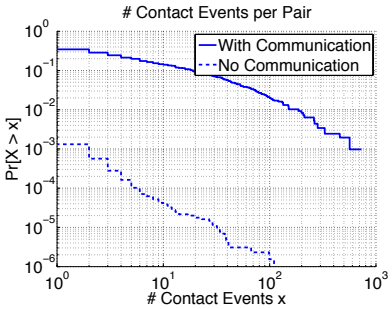


(a) CCDF of number of communication events for Facebook friends and Stumbl friends (log-log scale).

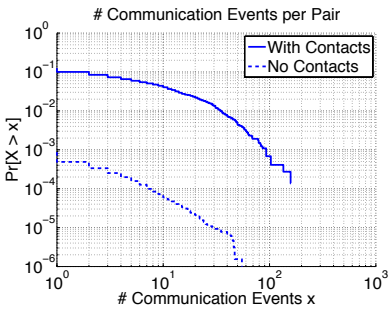


(b) Frequency of meetings vs. number of communication events.

Figure 10: Relation between meetings and communication events (Stumbl dataset).



(a) Impact of communication on number of contacts.



(b) Impact of mobility tie on number of communication events.

Figure 11: Relation of mobility and communication ties (Gowalla dataset).

number of communication events with general Facebook friends (including strong and weak mobility ties). Fig. 10a shows the complementary cumulative distribution functions of the pairwise number of communication events, for Stumbl friends, compared to Facebook friends. The plot shows that the number of communication events between Stumbl friends is indeed much higher than between "normal" Facebook friends. *In fact, on average a user communicates about* 10 *times more often with a Stumbl friend.* Yet, provided that a pair does meet face-to-face, the *frequency* of meetings is not a good indicator for the number of communication events. In Fig. 10b we see no correlation between the percentage of days a pair reports meetings and the number of communication events (only showing pairs that have at least one meeting and at least one communication event reported). Thus, while the information whether a pair does meet or not is a good predictor for Facebook communication, the number of meetings is not a good indicator for the intensity of communication.

Again, we confirm these results with the Gowalla dataset. In Fig. 11a, we plot the CCDF of the number of contacts for pairs that have at least one communication event and compare it to pairs that do not communicate. We observe that the distributions are qualitatively similar, but shifted by more than two orders of magnitude. Similarly, Fig. 11b shows that having a mobility tie increases the probability of communication by a factor greater than 100. Thus, communication largely happens between people who meet also face-to-face. For such local communication opportunistic networks are a viable solution which could manifest short message delivery delays.

These are rather preliminary results and the matter requires further research. However, already with the present data we can point out some implications. First, the finding that communication is more "local" than social connections is a strong argument in favor of opportunistic networks. In the future, more detailed analysis could provide answers to where opportunistic network are useful and in which cases infrastructure is required (i.e., for combined opportunistic and infrastructure networks). Second, in order to model data traffic in opportunistic networks, we should consider that pairs with strong mobility ties are more likely to communicate. Thus, realistic traffic models should be combined with realistic mobility models.

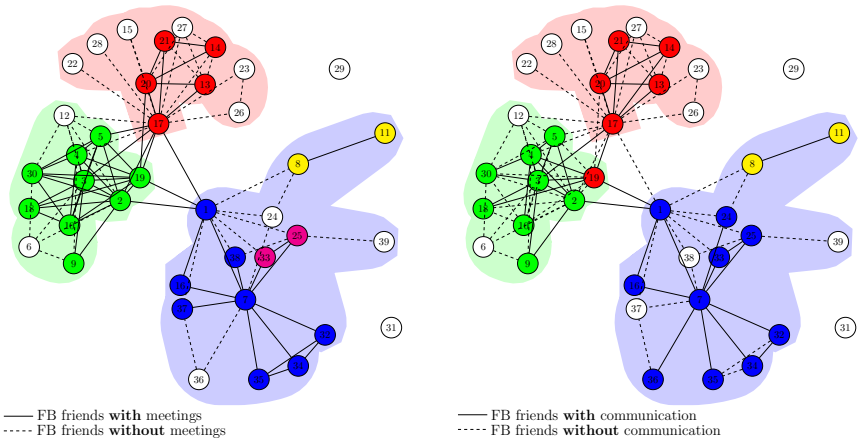
6. Multi-dimensional Network Analysis

In the last section we measured how the different dimensions of ties relate to each other *for individual pairs* (i.e., what is the probability of a pair having a tie in one relation type, given there is a tie in another relation type). In this section we go a step further and look at questions about macroscopic structure beyond pairs: *Are the communities that discuss with each other the same as the groups that meet each other? Are the central and influential people the same among different networks?* Etc. Answers to these questions are important for designing opportunistic protocols. One typical example is the class of social routing protocols (e.g., [5]) making decisions based on the social graph which are executed on the mobility graph (contacts).

To answer these questions about structure, we define graphs in all three relations: the social graph, the mobility graph and the communication graph on the set of Stumbl users and on the set of Gowalla users. Using tools and metrics from the field of complex network analysis we can then see how these graphs relate to each other.

6.1. Social, Meeting and Communication Graphs

We start with describing the three graphs and measuring some simple standard metrics of structure (avg. node degree, path lengths, clustering coefficient, etc.) [31] to gain an impression of the graphs' characteristics.



(a) Meeting graph.

(b) Communication graph.

Figure 12: Meeting and communication graphs. Communities are color coded: background colors = communities in the social graph; node colors = communities in the meeting, resp. communication graphs (Stumbl dataset).

6.1.1. Social Graph

Stumbl: We define the social graph $G^s_{soc}(V^s, E^s_{soc})$ such that the set of nodes V^s consists of the 39 Stumbl experiment participants[13] and the set of edges contains all pairs among them which are friends in Facebook. We have a total of 94 edges ($|E^s_{soc}| = 94$).

Gowalla: Similarly, we define the graph of social ties in Gowalla $G^g_{soc}(N^g, E^g_{soc})$, with $|V^g| = 978$ and $|E^g| = 1396$.

6.1.2. Meeting graph

Stumbl: For the meetings we define a *weighted graph* $G^s_{met}(V^s, E^s_{met}, W^s_{met})$ where we place an edge between pairs of nodes with at least one reported meeting. Further, the weights in the $|V^s| \times |V^s|$ weight matrix $W^s_{met} = \{w_{ij}\}$ indicate the strength of the mobility tie between each pair: w_{ij} is the percentage of days i and j have reported meetings (i.e., $w_{ij} = 0.5$ means that i and j met in half the days in which they reported data). Since both, i and j report meetings, we get two values w_{ij}. To make the matrix symmetric, we take the mean of both values for our analysis.

Note that our methodology of FB application based collection of meeting data allows only to learn about meetings between FB friends, not for accidental meetings between strangers or familiar strangers. Thus, the edges in the meeting graph are a subset of the edges of the social graph: $E^s_{met} \subset E^s_{soc}$.

[13]Note that we are not considering "external" nodes (non-participating nodes which get chosen as Stumbl friends by participants), since we have no information about relations between them.

The meeting graph is shown as a subset of the social graph in Figure 12a: solid line edges are nodes with meetings, whereas dashed edges are FB friends without meetings. We can see that the ties are not distributed randomly in the graph. Instead, there are gregarious nodes which are much better connected than the average nodes (e.g., node 17). We will discuss this property in Section 6.3. Further, we see (highlighted with colors) strongly connected communities, which we will discuss in more detail in Section 6.4.

Gowalla: For the Gowalla trace, we define the graph $G^g_{met}(V^g, E^g_{met}, W^g_{met})$. We use the number of contacts we observe in the Gowalla dataset as the pairwise mobility tie strength and define the matrix W^g_{met} as the matrix of weights with w_{ij} being the observed number of contacts between nodes i and j. A more detailed description and analysis about the Gowalla meeting graph (and contact graphs from other mobility traces) can be found in [32].

6.1.3. Communication graph

Stumbl: To define the communication graph $G^s_{com}(V^s, E^s_{com}, W^s_{com})$ we place an edge between pairs of nodes with at least one FB communication event (post, like, comment). Further, we weigh the tie strength by the number of communication events over the duration of the experiment in the weight matrix $W^s_{com} = w_{ij}$.

Again, with our methodology we can only collect communication events between FB friends, thus, $E^s_{com} \subset E^s_{soc}$. The communication graph is shown in Figure 12b. As for the meeting graph, we observe "hubs" of strongly connected nodes as well as tightly connected communities.

Gowalla: In order to infer how often two nodes communicate with each other, we account for mentions in the Twitter dataset. We have a total number of $\sim 37'000$ mentions among $3'787$ pairs of heavy users. For our analysis, we require an undirected tie strength. Hence, we define the communication graph $G^g_{com}(V^g, E^g_{com}, W^g_{com})$ where W^g_{com} is the matrix of weights with w_{ij} the sum of mentions of i to j and j to i.

6.2. Structural comparison

Due to the limited size of the Stumbl networks, we can look at them visually in Fig. 12. Comparison of the meeting graph and communication graph shows that they are highly similar – which is a surprising insight: it is not intuitive that the structure of who communicates with whom creates the same topology as who meets whom. In the following we compute some standard network analysis metrics (node degrees, path lengths, clustering coefficients) to confirm the structural similarity of the graphs in the Stumbl as well as the Gowalla data[14].

Table 2 summarizes standard graph metrics of all three graphs. The first thing to notice is that all graphs have a *giant component* [31], i.e., a connected component which spans the majority of nodes. In the social graph of the Stumbl data, this component covers all but 2 isolated nodes (i.e., 95% of the nodes are part of the giant component).

[14]Note that for the following measurements, we only take into account the binary graphs, since most of the metrics have clear and intuitive definitions for the binary case but not for the weighted case. We will use the weights mainly for the community analysis in Section 6.4.

	Stumbl			Gowalla		
	G^s_{soc}	G^s_{met}	G^s_{com}	G^g_{soc}	G^g_{met}	G^g_{com}
Size of largest component	95%	62%	62%	62%	76%	90%
Avg Degree $\bar{k}$	4.8	2.9	2.1	2.8	5.9	7.7
Clustering Coeff C	0.63	0.45	0.37	0.17	0.19	0.35
Norm. Clustering Coefficient C/C_r	5	5.8	6.9	59	32	44
Avg. Shortest Path Length $\bar{g}$	2.5	2.4	2.7	4.3	4.6	2.6
Norm. Avg. Shortest Path Length $\bar{g}/\bar{g_r}$	1.1	0.71	0.54	0.65	1.2	0.78

Table 2: Structural metrics of different networks.

For both, communication and meeting graphs, the giant component covers 62% of all nodes. By visual comparison we see in Figure 12 that for both graphs the giant components cover almost the same nodes: 22 of them are part of the giant component of both graphs. In the Gowalla data, the values are comparable, though in this case the communication graph has the largest giant component covering about 90% of all nodes.

Further, we are interested in the average node degree, i.e., the average number of neighbors of a node. In the Stumbl data, the nodes have an average of 2.9 neighbors in the meeting graph and an average 2.1 in the communication graph. Since in the Stumbl data these two graphs are subgraphs of the social graph, it is clear that they are less dense (have smaller average degree) than the social graph. In the Gowalla trace, where we also measure communication and meetings between nodes that do not have a social tie, the situation is different: here, the communication graph has the highest density (highest average node degree). This indicates that Twitter fosters communication even between people that do not meet face to face.

A typical property observed in a range of networks from different origins is high transitivity of relationships, manifesting itself in high clustering coefficients. The clustering coefficient of node u is defined as (e.g., [31])

$$C_u = \frac{\text{number of triangles connected to } u}{\text{number of triples connected to } u},$$

and the network clustering coefficient is the average of all node clustering coefficients $C = 1/|V| \sum_u C_u$. Since the expected random clustering coefficient depends on the density of the graph, we normalize the network clustering coefficient by $C_r = \bar{k}/|V|$, the corresponding random graph's expected clustering coefficient ($\bar{k}$ is the average degree over all nodes). Table 2 shows, that in all the Stumbl graphs the clustering coefficient is between 5 and almost 7 times higher than for a corresponding random graph. The Gowalla graphs are even more clustered, with values ranging from 32 to 59.

Another characteristic property of complex networks is the average path length. The average path length $\bar{g}$ is the shortest path averaged over all pairs of nodes, between which there exists a path. Again, the expected shortest path length for a random network $\bar{g_r} \approx ln(|N|)/ln(\bar{k})$ depends on the density, hence, we normalize the path length by $\bar{g_r}$ to make it comparable among the different graphs. We observe across all graphs that paths are short across the two datasets and all graphs. Note that short

	G^s_{soc}	G^s_{met}	G^s_{com}
G^s_{soc}	–	0.91	0.73
G^s_{met}	0.91	–	0.61
G^s_{com}	0.73	0.61	–

(a) Stumbl dataset

	G^g_{soc}	G^g_{met}	G^g_{com}
G^g_{soc}	–	0.36	0.43
G^g_{met}	0.36	–	0.47
G^g_{com}	0.43	0.47	–

(b) Gowalla dataset

Table 3: Spearman correlation coefficients for degree centrality ranking.

paths (compared to structured graphs like grids or rings) is a typical property of random graphs [31]. *Thus, our graphs show typical small world properties: Short average path length and high clustering coefficients.*

In conclusion, the macroscopic structure of the graphs is very similar. This is evident by visual comparison but also from simple standard structural metrics.

6.3. Central nodes

In the last subsection we have focused on network wide metrics to characterize the graphs. We now want to zoom in and look at individual nodes: *Are the sociable, well connected nodes the same across different networks, or do different nodes play the roles of "hubs"?*

In complex network analysis, the metric for "importance" of a node is called centrality. As there are different ways of defining importance – different problems (e.g., message routing, diffusion, resilience) call for different definitions of importance – there are different centrality metrics [33]. In the following, we use two commonly used centralities: *degree centrality* and *betweenness centrality*.

Degree centrality: Degree centrality simply counts a node's neighbors. With A being the adjacency matrix of a graph $G(V, E)$, i.e., $a_{ij} = 1$ if $\{i, j\} \in E$ and $a_{ij} = 0$ otherwise, the degree centrality of node u is

$$deg(u) = \sum_{v \in V} a_{uv}.$$

Depending on the process at hand, degree centrality captures how many peers a node can reach/infect/influence. For many applications (e.g., search [34]) it is hence beneficial to identify nodes with high degree centrality as the important "players" of the network.

To answer the question whether central nodes in one graph are also central in the other graph, we rank the nodes according to their degree centrality. For each relation type (social, meeting, communication) we obtain a vector, i.e., r_{soc}, r_{met} and r_{com} where the ith elements contains the degree centrality rank of node i in the respective graph. Using the *Spearman rank correlation coefficient* we can now compare these rankings. Table 3 shows the correlation coefficients for all combinations of graphs.

	G^s_{soc}	G^s_{met}	G^s_{com}
G^s_{soc}	–	0.89	0.93
G^s_{met}	0.89	–	0.83
G^s_{com}	0.93	0.83	–

(a) Stumbl dataset

	G^g_{soc}	G^g_{met}	G^g_{com}
G^g_{soc}	–	0.20	0.38
G^g_{met}	0.20	–	0.23
G^g_{com}	0.38	0.23	–

(b) Gowalla dataset

Table 4: Spearman correlation coefficients for betweenness centrality ranking.

We observe that the correlation coefficients are very high for the Stumbl data[15]. While this may not surprise (recall that G^s_{met} and G^s_{com} are subgraphs of G^s_{soc}) for correlation of meeting and social, resp. communication and social, the correlation of meeting and communication is very surprising: if we are able to identify the sociable nodes in terms of communication we get the hubs in terms of meetings as well – and vice versa. In the Gowalla data, the correlation is a little less pronounced but still quite strong.

Betweenness centrality: For other processes it is less important how many neighbors a node can reach directly, but rather how often it falls on a shortest path between two other nodes. This measures how important the node is in terms of relaying/intercepting/controlling shortest path communication in the network. To quantify this, betweenness centrality is defined as

$$bet(u) = \sum_{i \in V} \sum_{j \in V} \frac{g_{ij}(u)}{g_{ij}}.$$

$g_{ij}(u)$ counts the number of shortest paths between i and j of which u is a part of, and $g_{i,j}$ counts all shortest paths between i and j.

Table 4 shows the rank correlation coefficients for the betweenness centrality ranking. As for degree centrality ranking, we find that the correlation is extremely high in the Stumbl data. Again, in the Gowalla data the correlations are less strong.

In conclusion, we found that the importance of a node in one graph (i.e., one relation type) can tell us a lot about its importance in another graph. If we can identify central nodes in one dimension this is a strong predictor of high centrality in another dimension, in particular in smaller scenarios like we observe it in Stumbl.

6.4. Communities

Besides node centrality, community structure [35] is another property often observed in different social networks: Nodes tend to group in clusters which are strongly connected internally, but have only weak connections to the outside. The existence of strong communities has various implications for opportunistic networks: On one hand, it implies high potential for node cooperation and community-based trust mechanisms.

[15]The p-values (not shown) are very small, i.e., much smaller than 0.05, indicating that the correlations are statistically relevant.

On the other hand, it may also imply high convergence times for processes running over the graph, since there may be strong bottlenecks between communities.

To detect communities in the contact graph, we apply the widely used Louvain community detection algorithm [36]. To quantify the modularity of the resulting node partitioning, we apply the commonly used modularity function [37]:

$$Q = \frac{1}{2m} \sum_{ij} \left(w_{ij} - \frac{d_i d_j}{2m} \right) \delta(c_i, c_j),$$

where $d_i = \sum_j w_{ij}$ is the strength (the sum of a node's weights) of node i and $m = \frac{1}{2} \sum_j d_j$ is the total weight in the network[16]. c_i denotes the community of node i. Hence, the Kronecker delta function $\delta(c_i, c_j)$ is one if nodes i and j are members of the same community and zero otherwise. $Q = 0$ is the expected quality of a random community assignment and modularities of above $Q = 0.3$ are typically reported for networks of various origins (social, biological, etc.). The values we obtain for our graphs are reported in the following table:

G^s_{soc}	G^s_{met}	G^s_{com}
0.54	0.53	0.62

(a) Stumbl dataset

G^g_{soc}	G^g_{met}	G^g_{com}
0.63	0.84	0.89

(b) Gowalla dataset

Table 5: Modularity values of community assignments.

Modularity is very high in all cases, especially in the Gowalla data. In Figures 12, the nodes of the Stumbl graphs are colored according to the communities they are assigned to in the meeting/communication graph and the background is shaded to indicate community membership in the social graph. We observe that for the nodes not isolated in the meeting/communication graph, the community assignments are almost identical in all three graphs.

Summarizing, in terms of *structural properties*, *central nodes*, as well as *community assignment*, we have found very strong similarity across the social, meeting and community graphs.

7. Conclusion

We have presented an analysis of two datasets of self-reported data about social, mobility and communication ties of online social network users. The first dataset is collected using Stumbl, a Facebook application to collect data for opportunistic networking research. Stumbl automatically collects interaction events using the Facebook API and relies on user reports about the type of their social relationships and the face-to-face meetings. The second dataset is gathered from publicly available data from Gowalla and Twitter.

[16]For the binary social graph, all weights are 1.

The analysis of the two datasets has revealed that all three dimensions of tie strength depend on each other. (1) Social ties and mobility ties are tightly coupled. In the Gowalla dataset, we observe that having a social tie increases the probability of having a social tie by at least two orders of magnitude. Further, using the Stumbl data, we have shown that the type of social tie (friend, family, colleague or acquaintance) has strong impact on context, duration and frequency of meetings. Consequently, we argue that having this information is valuable for instance for opportunistic routing protocols. (2) Similarly, communication and mobility are correlated. In the Gowalla dataset, we see that having an edge in one dimension increases the probability of having an edge in the other dimension by a factor larger than 100. Further, using the Stumbl data, we can show that the number of Facebook communication events differs for different relationship ties, a fact which should be considered when modeling traffic in opportunistic network. (3) And finally, social ties and communication are strongly related, both in the Gowalla and Stumbl data. People communicate preferentially with friends they also have face-to-face meetings. Thus, communication ties are more local than social ties.

Further, we have used metrics from complex network analysis to show that the structure of the social, meeting and communication graphs manifest very similar structure. Not only do they all show small-world properties (short average path lengths and high clustering), but the nodes play similar roles across tie type: the hubs of one dimension are also hubs in the other dimension, and the communities of one dimension are also communities in the other dimension.

In the future, we plan to run bigger Stumbl experiments with more participants, in order to get the level of detail Stumbl provides for a larger number of nodes. The challenge is to provide incentives to the users to regularly report true data about their face-to-face meetings. Using game mechanisms, if designed carefully, could be a promising approach to spread the application.

Acknowledgment

This work was partially funded by the European Commission under the SCAMPI (258414) FIRE Project.

References

[1] The Aka Aki Network. `http://www.aka-aki.com/`. [Online]. Available: http://www.aka-aki.com/

[2] B. Distl, G. Csucs, S. Trifunovic, F. Legendre, and C. Anastasiades, "Extending the reach of online social networks to opportunistic networks with PodNet," in *MobiOpp*, 2010.

[3] A.-K. Pietiläinen, E. Oliver, J. LeBrun, G. Varghese, and C. Diot, "Mobiclique: Middleware for mobile social networking," in *WOSN*, 2009.

[4] T. Spyropoulos, K. Psounis, and C. S. Raghavendra, "Spray and wait: an efficient routing scheme for intermittently connected mobile networks," in *WDTN*, 2005.

[5] A. Mtibaa, M. May, C. Diot, and M. Ammar, "PeopleRank: Social opportunistic forwarding," in *IEEE INFOCOM*, 2010.

[6] C. Boldrini, M. Conti, and A. Passarella, "Contentplace: social-aware data dissemination in opportunistic networks," in *ACM MSWiM*, 2008.

[7] P. Hui and J. Crowcroft, "How small labels create big improvements," in *PerCom Workshops*, 2007.

[8] P. Hui, J. Crowcroft, and E. Yoneki, "BUBBLE Rap: Social-based Forwarding in Delay Tolerant Networks," in *ACM MobiHoc*, 2008.

[9] T. Hossmann, T. Spyropoulos, and F. Legendre, "Know thy neighbor: Towards optimal mapping of contacts to social graphs for DTN routing," in *IEEE INFOCOM*, 2010.

[10] X. Zhang, J. Kurose, B. N. Levine, D. Towsley, and H. Zhang, "Study of a Bus-Based Disruption Tolerant Network: Mobility Modeling and Impact on Routing," in *Mobicom*, 2007.

[11] A. Lindgren, A. Doria, J. Lindblom, and M. Ek, "Networking in the land of northern lights: two years of experiences from dtn system deployments," in *WiNS-DR*, 2008.

[12] T. Henderson, D. Kotz, and I. Abyzov, "The changing usage of a mature campus-wide wireless network," in *ACM MOBICOM*, 2004.

[13] A. Mtibaa, A. Chaintreau, J. LeBrun, E. Oliver, A.-K. Pietilainen, and C. Diot, "Are you moved by your social network application?" in *WOSN*, 2008.

[14] P. Hui, A. Chaintreau, J. Scott, R. Gass, J. Crowcroft, and C. Diot, "Pocket switched networks and human mobility in conference environments," in *WDTN*, 2005.

[15] N. Eagle, A. S. Pentland, and D. Lazer, "Inferring friendship network structure by using mobile phone data," *PNAS*, 2009.

[16] V. Lenders, J. Wagner, and M. May, "Measurements from an 802.11b mobile ad hoc network," in *IEEE EXPONWIRELESS*, 2006.

[17] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott, "Impact of human mobility on the design of opportunistic forwarding algorithms," in *IEEE INFOCOM*, 2006.

[18] T. Karagiannis, J.-Y. Le Boudec, and M. Vojnovic, "Power law and exponential decay of inter contact times between mobile devices," in *ACM MobiCom*, 2007.

[19] V. Conan, J. Leguay, and T. Friedman, "Characterizing pairwise inter-contact patterns in delay tolerant networks," in *ACM Autonomics*, 2007.

[20] W. J. Hsu, T. Spyropoulos, K. Psounis, and A. Helmy, "Modeling time-variant user mobility in wireless mobile networks," in *IEEE INFOCOM*, 2007.

[21] C. Tuduce and T. Gross, "A mobility model based on WLAN traces and its validation," in *IEEE INFOCOM*, 2005.

[22] M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi, "Understanding individual human mobility patterns," *Nature*, 2008.

[23] C. Song, Z. Qu, N. Blumm, and A.-L. Barabási, "Limits of predictability in human mobility," *Science*, 2010.

[24] D. Wang, D. Pedreschi, C. Song, F. Giannotti, and A.-L. Barabási, "Human mobility, social ties, and link prediction," in *KDD*, 2011.

[25] d. boyd, *None of this is Real*, New York, 2008.

[26] S. Scellato, A. Noulas, and C. Mascolo, "Exploiting place features in link prediction on location-based social networks," in *KDD*, 2011.

[27] T. Hossmann, T. Spyropoulos, and F. Legendre, "A complex network analysis of human mobility," in *NetSciCom*, 2011.

[28] T. Hossmann, F. Legendre, G. Nomikos, and T. Spyropoulos, "Stumbl: Using facebook to collect rich datasets for opportunistic networking research," in *AOC*, 2011.

[29] G. Nomikos, "Studying social-driven mobility: Comparing face-to-face meetings to online social network activity," Master's thesis, ETH Zurich, 2010.

[30] P. V. Marsden and K. E. Campbell, "Measuring tie strength," *Social Forces*, 1984.

[31] M. E. J. Newman, "The structure and function of complex networks," March 2003.

[32] T. Hossmann, T. Spyropoulos, and F. Legendre, "Putting contacts into context: Mobility modeling beyond inter-contact times," in *MobiHoc*, 2011.

[33] L. C. Freeman, "Centrality in social networks: conceptual clarification," *Social Networks*, 1979.

[34] L. A. Adamic, R. M. Lukose, A. R. Puniyani, and B. A. Huberman, "Search in Power-Law Networks," *PHYS.REV.E*, 2001.

[35] S. Fortunato, "Community detection in graphs," *Physics Reports*, 2010.

[36] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *J.STAT.MECH.*, 2008.

[37] M. E. J. Newman, "Modularity and community structure in networks," *PNAS*, 2006.